10-2-2018

The Gender Gap in Stem Attainment

Tim R. Sass
Georgia State University, tsass@gsu.edu

W. J. Usery Workplace Research Group Paper Series

Working Paper 2018-10-2
October 2018

The Gender Gap in Stem Attainment

Tim R. Sass
Georgia State University

Georgia State University
Andrew Young School
of Policy Studies

THE GENDER GAP IN STEM ATTAINMENT *

Tim R. Sass
Department of Economics
Georgia State University

October 24, 2018

Abstract

I investigate the determinants of high school completion and college attendance, the likelihood of taking science, technology, engineering or math (STEM) courses in the first year of college and the probability of earning a degree in a STEM field. The focus is on women, who tend to be underrepresented in STEM fields. Tracking four cohorts of students throughout Florida, women perform nearly as well as men on math achievement tests through high school and are more likely to finish high school and attend college than males. Among college students, however, women are less likely than men to take courses in the physical sciences in their first year and are less likely to earn a degree in physics or engineering, even after adjusting for pre-college test scores. Gender matching of students and math/science teachers in middle and high school tends to increase the likelihood that female college freshman will take at least one STEM course. However, conditional on first-year coursework, neither gender matching at the secondary or college levels appears to have any effect on the likelihood of completing a major in a STEM field. For all students, having high school math and physics teachers with a degree in math or physics, respectively, (as opposed to education) is associated with a higher likelihood of taking STEM courses as college freshmen.

---

*I wish to thank the staff of the Florida Department of Education's Education Data Warehouse for their help in obtaining and interpreting the data used in this study. Funding for this research was provided by the Gates and Smith-Richardson Foundations as part of a grant to the Center for the Analysis of Longitudinal Data in Education Research (CALDER). Katherine Key provided excellent research assistance. Any opinions or errors are solely attributable to me, however.

I. Introduction

There is growing concern that the United States does not produce a sufficient number of students majoring in science, technology, engineering and math (STEM) fields in order to remain globally competitive. Of particular concern is the underrepresentation of women in STEM fields. Various hypotheses have been forwarded to explain why there are relatively few women in STEM areas, including negative peer effects of male students in math and science courses and lack of female instructors as role models.

Most of the existing evidence on enrollment and persistence in STEM majors is based on experiences while in college. However, this may obscure important influences of student preparation and experiences in high school or earlier. The focus on courses, instructors and grades while in college is primarily a matter of data availability. Most extant studies have relied on college transcript data, which provide little information about student experiences prior to attending college.

The present study of STEM degree attainment is the first to track individual students from elementary school through the end of college and the first to link individual students to their STEM instructors in middle school, high school and college. It is also the first study to simultaneously consider the impacts of both high school instructor demographics and the training of high school teachers on their students' choices of coursework and major in college. The availability of linked K-12 and post-secondary transcript data allow one to determine where gaps emerge between males and females and the relative importance of pre-college and within-college experiences in determining STEM educational attainment. Such information is important in determining where to target interventions designed to promote female participation in STEM fields.

The evidence reveals that women perform nearly as well as men on math achievement tests through high school and are more likely to finish high school and attend college than males. Among college students, however, women are much less likely than men to earn a degree in a STEM field, even after adjusting for pre-college test scores. There is some evidence that gender matching of students and secondary teachers increases the likelihood that women take STEM classes as college freshmen. Likewise, taking high school high school courses from teachers who possess a baccalaureate degree in math or science (rather than education) is associated with a higher probability of taking STEM courses as a college freshman. However, conditional on first-year college coursework, gender matching of students and college instructors is not associated with completing a degree in a STEM field.

II. Existing Evidence

Spurred by the intense policy interest in the underrepresentation of women in STEM fields, there has been a recent surge in research on the determinants of entry and persistence in STEM majors. Most of this recent work focuses on major choice conditional upon enrolling in college.

A. STEM Persistence Among College Students

Bettinger and Long (2005) employ individual level data on students who entered four-year public colleges and universities in Ohio in 1998 and 1999. They find that exposure to a female college instructor in their first college course in a subject has mixed effects on female students' participation in STEM. Female students who initially had a female instructor were less likely to take additional courses in biology and physics than similar students whose first professor in the subject was male. However, initial exposure to a female instructor boosted the likelihood that a female student would take an additional course in the subject and the total number of hours of credit hours in geology, mathematics and statistics. They found no effects of instructor gender on the major choices of female college students in STEM fields. In later work, Bettinger (2010), finds STEM fields only retain about one-half of students. Further, the proportion of first-year courses taken in STEM fields is directly correlated with an eventual major in STEM.

Price (2010) analyzes the same Ohio data, but for a longer time period, 1998-2002. He focuses on the impact of race and gender matching of students and instructors in first-semester freshman STEM courses on student persistence in STEM majors. As in Bettinger and Long (2005), he finds that female students are less likely to persist as the proportion of their STEM courses taught by female instructors rises.

Hoffmann and Oreopoulos (2009) use data from the University of Toronto to analyze the impact of instructor gender in large first-year undergraduate courses on persistence within a course and later course taking in the same subject. With multiple subjects, they are able to employ student fixed effects to control for unobserved student heterogeneity. In the fixed-effects models, they find that gender matching of students and instructors has no effect on the likelihood of dropping the course, but does increase the number of additional courses taken in the subject. This later effect does not hold for math and science courses, however, where gender matching actually reduces the number of additional courses taken in the subject.

Kokkelenberg and Sinha (2010) analyze course grades and graduation probabilities of students at SUNY-Binghamton. They find that student SAT math scores and having AP credits in a STEM field are positively correlated with completing a major in a STEM field. However, conditional on AP credits and SAT scores, women are less likely than men to graduate with a major in a STEM field. Peer influences were mixed; increases in the proportion of female students increased the likelihood a woman would earn an “A” in sophomore math courses, but decreased the probability of obtaining a grade of “A” in junior-level math courses and had no significant effect on grades earned in biology courses. Similarly, having a female instructor in a college math or biology course did not improve the chances a female student would earn an “A” and was negatively correlated with the probability of a woman earning an “A” in introductory biology.

Like Bettinger and Long (2005), Griffith (2010) finds a zero or negative correlation between percent of female faculty in STEM and female student persistence in a STEM major. However, the proportion of STEM PhDs awarded to women at a university is positively correlated with persistence in a STEM major into the sophomore year, but not into the senior year.

Carrell, Page and West (2010) analyze the effects of instructor gender on student achievement, course taking and STEM major choice at the U.S. Air Force Academy. The Air Force Academy requires all students to take a set of standardized introductory courses and randomly assigns students to classrooms, thereby avoiding any possible selection bias associated with student course and instructor choice. In contrast to other studies, they find substantial positive effects of female instructors on female student’s performance in math and science courses at the Academy. Further, high-performing young women, as measured by SAT math scores, are more likely to take additional STEM courses and graduate with a degree in a STEM major if they have introductory courses from a female instructor. For these high-ability women taking introductory courses from a female instructor, the STEM gender gap is eliminated. It is unclear, however, whether their strong gender matching findings are due to their rigorous experimental design or are related to unique aspects of the female students attending an elite military academy.

B. Pre-College Influences on Educational Attainment in STEM Fields

The literature relating pre-college experiences to college outcomes is also thin. Dee (2007), employing a cross-subject student fixed effects strategy, finds that exposure of 8^{th} grade female students to a female science teacher had insignificant effects on science test scores and assignment to a female math teacher significantly lowered math achievement for girls. However, Dee's evidence ends at the middle school level and provides no direct evidence on outcomes in high school and college.

Park, Behrman and Choi (2018) exploit the random assignment of students to high schools in Korea to estimate the causal effects of same-sex schools on math college entrance exam scores, student expectations of college attendance and major choice while in high school and actual college attendance and major choice two years after high school. They find that attending an all-girls school has no statistically significant effect on general math test scores, math-science test scores, Korean test scores or English test scores. Further, attending an all-girls high school has no significant effect on interest in science, student expectations or actual choice of a STEM major.

Anelli and Peri (2017) link data on graduates of college-preparatory high schools in Milan to enrollment records of universities in Milan in order to study the effects of high school gender composition on college major choice, academic performance in college and labor market outcomes. In the Italian system, high school students are randomly grouped into "classes" during their freshman year and maintain these groupings throughout high school. Rather than study specific STEM fields, they divide majors into "prevalently male" (PM) and "prevalently female" (PF) categories and consider the determinants of initial major choice. They find no evidence that the share of own-gender peers affects in high school impact the probability of choosing PM or PF majors for either males of females.

Card and Payne (2017) analyze data from a cohort of high school students in Ontario, Canada to determine the influence of choices made in high school on entry into STEM programs in college. Two unique aspects of the Canadian system lead them to focus on course selection and grades earned in high school. First, unlike for most U.S. colleges and universities, high school graduates in Canada apply to specific programs at universities in a province, rather than simply admission to the institution. Second, in nearly all cases, admission decisions are based solely on applicant's grades in standardized courses taken in the fourth year of high school. Card and Payne find that differences in high school course taking explain only a small proportion of the gender gap in STEM program entry. Rather, the gap is mainly due to higher rates of college going by women, which means that a smaller proportion of female college entrants possess the pre-requisites for STEM programs in college.

Two recent studies link high school and college data in the U.S., Bottia, et al. (2015) focuses on the relationship between the gender of high school faculty and the likelihood that young women declare a major and complete a major in a STEM field. The study employs data on a single cohort of 12,550 students who graduated from North Carolina public schools in 2004 and enrolled in a public 4-year college or university the same year. Student records from grades 7-12 are linked to post-secondary records, which provide information on the courses taken by individual students in high school as well as school-level measures of faculty composition in high school. Their data do not link students and teachers to individual classrooms at either the high school or post-secondary level, however, so it is not possible to determine the gender composition of faculty that individual students are exposed to. They find that higher proportions of female STEM faculty at the high school level are associated with greater likelihoods that women declare a STEM major and graduate with a STEM degree. These effects are greatest for female students with the strongest math skills.

Like Bottia et al. (2015), Shi (2018) utilizes individual-level data on public school students in North Carolina who matriculate into the University of North Carolina (UNC) system of colleges and universities. Her data include multiple cohorts of students that were enrolled in a UNC system school during the period 2003-2008, nearly 140,000 students in all. Linked high school test-score records provide evidence on math and science courses taken and the level of performance on end-of-course exams. Her analysis focuses on the likelihood that women majored in engineering, rather than the broader mix of STEM fields. She analyzes both a student's "first choice major" at the time of taking the SAT (typically in their junior or senior year of high school), major choice after having completing 30 credit hours (approximately one year of study in college) and attrition from engineering for those who declare an engineering major as freshmen. Her results indicate that women are 12 percent less likely than men to have declared major in engineering after 30 credit hours. Conditioning on high school course taking, performance on high school end-of-course exams, SAT scores and high school fixed effects only lowers the gap to 11 percentage points. Women persist at comparable rates to men once they declare a major in engineering, however.

III. Data

The data for this study come from a variety of sources. The primary source for student-level information is the Florida Department of Education's K-20 Education Data Warehouse (K-20 EDW), an integrated longitudinal database covering all public school students and teachers in the state of Florida. For K-12 students, the K-20 EDW provides demographic information, enrollment and attendance, program participation, disciplinary actions and achievement test scores, beginning in 1995. Florida began testing students statewide in 1997/98, with the introduction of the "Sunshine State Standards" Florida Comprehensive Achievement Test (FCAT-SSS). The FCAT-SSS is a criterion-based exam designed to test for the skills that students are expected to master at each grade level. It is a "high-stakes" test used to determine school grades, student retention in some grades and passage of the 10th grade exam was a requirement for graduation from high school for many years. The FCAT-SSS was initially administered to selected grades but was later expanded to grades 3-10 in 2000/01. Beginning in 1999/2000, a second test, the FCAT Norm-Referenced Test (FCAT-NRT) was added in each of grades 3-10. The FCAT-NRT, was a custom form of the Stanford Achievement Test used throughout the country. No accountability measures were tied to student performance on the NRT. Florida stopped administering the FCAT-NRT after 2007/08. The FCAT-SSS exam was replaced with the FCAT 2.0 beginning in 2010/11.

As the name implies, the K-20 EDW also includes records for students enrolled in community colleges or four-year public universities in Florida. The K-20 EDW also contains information on the Florida Resident Assistance Grant (FRAG), a grant available to Florida residents who attend private colleges and universities in the state. Data from the National Student Clearinghouse (NSC), a national database that includes enrollment data from 3,300 colleges throughout the United States, is used to track college attendance outside the state of Florida, as well as any private college enrollment in Florida that the FRAG data do not pick up. Unfortunately, the Florida Department of Education's data-sharing agreement with the NSC expired in the latter part of the 2000s, so I can only reliably track students who attended private colleges and universities within Florida or any postsecondary institution outside of Florida through school year 2006–2007.[1] Enrollment, coursework and degree attainment information are available for all post-secondary students at public institutions in Florida. In addition, demographic information on post-secondary instructors is available as well.

---

[1] Information on the NSC is available at www.studentclearinghouse.org.

High school graduation is determined by withdrawal information and student award data from the K-20 EDW. While various diploma options exist, including a GED and a special-education diploma, I focus on receipt of a regular high school diploma. Students who withdrew with no intention of returning or exited for other reasons such as non-attendance, court action, joining the military, marriage, pregnancy, and medical problems, but did not later graduate, are counted as dropouts. It is not possible to directly determine the graduation status of students who leave the Florida public school system to attend a home-schooling program, to enroll in a private school or who move out of state.

The analysis sample covers four cohorts of 5th-grade students. Statewide achievement testing for 5th-grade students began in the 1997/98 school year, so the first cohort in the sample are students who attended 5th grade for the first time in 1997/98 and took the FCAT-SSS math exam. The final cohort is composed of students who were enrolled in 5th grade for the first time in 2000/01. Descriptive statistics for these four cohorts of students are provided in Table 1.

The last year for which I was able to obtain student data is 2012/13. Given that it takes at least three years to progress through middle school and high school completion typically takes four years, this means that each of the four cohorts can be tracked through high school and into the beginning of college. If I allow five years for college completion, all four cohorts can also be tracked through the end of college so long as I restrict the sample to students who do not repeat any grades in middle school. Descriptive statistics for this restricted sample are provided in Table 2. As one would expect, students who follow the normal progression through middle school have much higher test scores than those who repeat a grade or drop out before completing high school and thus the test score means in Table 2 are higher than those in Table 1.

IV. Analysis and Results

A. Descriptive Analysis of Pre-College Outcomes and College Attendance

In order to understand where the "leakages" in the STEM pipeline occur, I begin with a descriptive analysis of achievement differences in math prior to high school entry. Table 3 reports mean test scores in elementary and middle school by gender and race. While there are stark differences across racial/ethnic groups, within-race gender differences in math achievement are relatively small or non-existent. White girls tend to start out below white boys, with test scores that are 0.03 to 0.07 standard deviations lower in 3^{rd} grade. By 8^{th} grade, however, their math test scores are essentially equal (a 0.01 difference in favor of girls on the FCAT-SSS exam and a 0.02 advantage for boys on the FCAT-NRT exam). For Asians, boys enjoy a 0.03 to 0.04 standard deviation advantage in math achievement that remains fairly constant through 8^{th} grade. Among Hispanics, girls start out about 0.02 standard deviations lower in third grade, but their math achievement level equals or exceeds that of Hispanic boys by eighth grade, depending on the test. Counter to the trends for other racial and ethnic groups, black girls outscore their male counterparts by 0.10 to 0.13 standard deviations on third grade math exams and the gap remains fairly constant throughout middle school.

The high school experiences of students, broken down by race and gender, play out much like the achievement measures in elementary and middle school. Table 4 provides information on exit propensities by race/ethnicity and gender. As with test scores prior to high school, the differences in drop-out rates between white boys and girls are relatively small, with females about 1 percentage point less likely to drop out of high school. Gender differences in drop out probabilities are about twice as high for blacks and Hispanics, with young black and Hispanic men each about two percentage points more likely to drop out than their female counterparts.

Even if students complete five years of high school, they may not earn a regular high school diploma. They could receive a GED or (if they are a special education student) a special diploma or certificate of completion. Alternatively, they could remain enrolled, but still not have obtained a diploma within five years. As demonstrated in Table 5, within each racial/ethnic category, women are more likely to earn a standard high school diploma than are men. The male-female graduation differential is highest among blacks, with young black women nearly 10 percentage points more likely than young black men to earn a standard high school diploma within five years of starting high school.

A final measure of high school performance is achievement test scores. Table 6 provides average math test scores broken down by race/ethnicity and gender. Within racial/ethnic groups, the gender achievement differentials mostly remain relatively constant through high school. The one exception is for black students, where the math achievement advantage for young women is cut in half by 10th grade. This is likely due to the higher rate of drop out among black males in the early high school grades.

Conditional on earning a regular high school diploma, there are substantial differences in post-secondary educational choices between men and women. As reported in Table 7, within each racial/ethnic category, women are more likely to attend college than are men. The within-race gender differences are starkest for blacks, where young black women are nearly 20 percentage points more likely to enroll in either a two-year or four-year post-secondary institution.

Taken together, the descriptive evidence on test scores and educational attainment prior to college suggest that women would be as likely as men to complete a STEM major in college. I turn now to analyses of course selection, persistence and major choice among those students who make it to college.

B. Determinants of Initial College Coursework

Table 8 presents probit estimates of the probability of taking one or more STEM courses in the first year of college, conditional upon having earned a regular high school diploma within five years of starting grade 9 and enrolling in a four-year public university within a year of receiving their high school diploma. The first column reports estimates with only student gender and race/ethnicity in the model. Without any other controls, women are about 4 percentage points less likely than men to take at least one STEM course in their freshman year.[2] Estimates reported in the second column indicate that controlling for family income (5th-grade lunch status) and K-12 math test scores does not substantially alter this finding.[3]

As shown in Table 9, the relationships between STEM course taking and gender vary across science disciplines. Women are less likely than men to take any science course and less likely to take computer science, engineering, math or physics courses during their freshman year. They are more likely than men, however, to take at least one biology course or at least one statistics course during their first year of college.

Table 10 presents estimates of the determinants of first-year STEM course taking, controlling for the characteristics of middle and high school teachers and peers.[4] The first two specifications exclude high school fixed effects whereas the third and fourth specifications include them and thereby reflect within-high-school comparisons. Due to the incidental parameters problem, probit models with fixed effects may be biased (Greene, 2004). Therefore, I estimate

---

[2] Remedial math courses are excluded.

[3] Much of the influences of income and prior achievement are likely implicitly controlled for by restricting the sample to high school completers who immediately enter a four-year public university in Florida.

[4] Note that I control for pre-high school math test scores, but not scores in grades 9 and 10, so the high school teacher effects could be working through effects on high school math achievement.

linear probability models for the fixed-effects specifications. I present estimates from both probit and linear probability models without fixed effects for comparison purposes.

The advantage of including high school fixed effects is that they eliminate bias caused by unmeasured time-invariant school characteristics that are correlated with both high school teacher characteristics and with subsequent college course taking decisions. For example, suppose that schools in relatively affluent neighborhoods tend to have teachers with degrees in their subject area and parents in these neighborhoods tend to push their children toward STEM majors in college. What would appear to be an effect of teacher credentials could in fact be caused by unmeasured neighborhood characteristics. Including high school fixed effects would eliminate such potential biases.

The disadvantage of using high school fixed effects is that they may soak up important cross-school variation that would otherwise be captured by the variables of interest. For example, if gender matching of students and teachers in high school promotes later college STEM course taking, we would expect that female students from a school with a predominately female math/science faculty would fare better than would female students attending a school where the math and science faculty are mostly male. With high school fixed effects, these cross-school differences would be absorbed into the fixed effects, however. Put differently, identification of the effects of faculty identity come solely from within-school changes over time in faculty composition. Given a student typically attends high school for four years and we have only four cohorts of students, within-school variability may be somewhat limited.

For both the probit and linear probability the models without fixed effects, I find very similar positive student-teacher match effects for women. The gender gap in first-year STEM course taking (estimated to be -0.0325 in Table 9) is essentially eliminated if half of female students' middle and high school math and science courses are taught by women. The preparation of math and science teachers matters as well, however. Having at least one biology, chemistry or math course in high school taught by a teacher with a bachelor's degree in the relevant subject (rather than a math or science education degree or a degree in another science field) has small but statistically significant effects on the likelihood a student later takes at least one STEM course as a freshman in college. Likely due to the small number of high school instructors with degrees in physics, the estimated impact of having a high school physics teacher who majored in physics is imprecise and not statistically significant at conventional confidence levels. In contrast to the apparent influences of teachers, having a greater proportion of female students in math and science courses does not appear to boost the likelihood that a young woman will take STEM courses in their first year in college. In fact, the correlation between the fraction of female students in middle and high school math and science courses and STEM course taking during the first year in college is negative.

Many of the effects of high school teachers on college course taking decisions disappear when high school fixed effects are included in the model. As shown in the third column of Table 10, the positive match effects for women are eliminated when high school fixed effects are employed. Likewise, all of the positive teacher training effects go away as well. The only statistically significant effect that remains is a small negative effect on later STEM course taking of having a high school biology course taught by a teacher who majored in biology. In attempt to enhance inter-temporal variability (which is required for identification in the fixed effects model) I limit the sample to large high schools that employed 100 or more unique teachers. Results are presented in the fourth column of Table 10. The female indicator is now negative and statistically significant. In addition, the effect of having a high school chemistry course taught by a teacher with a degree in chemistry is now positive and statistically significant.

As an alternative to the fixed effects models, I also estimate a probit model that deals with potential endogeneity by employing instrumental variables. The advantage of this approach is it allows cross-school variation to be considered while still addressing potential bias from self-selection of high school instructors. Following the strategy employed by Bettinger and Long (2005) and Price (2010), I use average faculty composition variables as instruments for the characteristics of teachers a student has in high school.[5] Thus, for example, the school-wide proportion of biology teachers who possess a degree in biology during a student's high school career serves as an instrument for the indicator that a student was taught biology by a teacher who majored in biology in college. Results from the probit IV estimation are presented in the last column of Table 10. The IV results are qualitatively similar to those from the probit model without high school fixed effects. Gender matching of students and teachers in middle and high school math and science courses increases the likelihood of women taking at least one STEM course as a college freshman, an increase from 10 percent to 20 percent in the proportion of a young woman's middle and high school math and science courses taught by female teachers would translate into at least a 14 percentage point increase in the probability of taking one or more STEM courses during their first year in college (0.1 x 1.40). The likelihood of a female student taking one or more STEM classes in their first year in college is also enhanced if they had a high school math or physics course taught be someone who majored in the relevant subject in college. In contrast, it

[5] I calculate school average faculty composition for each student by taking the simple average of school/year/grade faculty composition over each of the school/grade/year combinations for the years each student was enrolled in grades 9-12.

appears that having more female students in middle and high school math and science courses actually diminishes the likelihood of taking a STEM course in the first year of college; a 20 percent increase in the proportion of classroom peers who are women is associated with a 10 percentage point reduction in the probability a female student will take one or more STEM courses during their first year in college.

C. Determinants of Completing a College Degree in a STEM Major

Table 11 presents results of estimating probit equations which predict completion of college majors (conditional on attending a Florida public university immediately after earning a high school degree), with and without controls for pre-college family income and student achievement test scores. The estimates from the model without pre-college controls reveal the expected pattern; women are less likely than men to successfully complete a major in a STEM field. This pattern generally holds across specific STEM majors. The one notable exception is biology and other life sciences, where women have a higher likelihood of earning a bachelor's degree than men.

The second panel of Table 11 presents estimates of STEM major completion in which family income (proxied by free/reduced-price lunch status in grade 5) and middle and high school math achievement scores are included as controls. Holding constant family income and prior test scores, the magnitude of the gender differences in reduced by one-half or more, though the general pattern still holds. Holding constant family income and individual math skills, women are more likely than men to complete a major in life sciences, but are less likely than men to earn a degree in engineering, math or physics.

The finding that women are less likely to earn bachelor's degrees in math, physical sciences and engineering than their male counterparts with equivalent resources and math skills begs the question of whether changes in faculty gender composition would likely alter the outcome. In order to gauge whether student-teacher matching by gender effects successful completion of a STEM major I estimate models of degree completion which include controls for both middle/high school and first-year-in-college matching of students and instructors, conditional on first-year coursework in college.

Estimates of the determinants of major completion are presented in Table 12. Not surprisingly, freshman-year coursework is strongly related to eventual degree attainment. The greater the number of freshman-year engineering courses taken the more likely a student will eventually earn an engineering degree. The same is true for math courses and math degrees and freshman science courses and degrees in chemistry, physics and life sciences. Controlling for pre-college influences and course selection in the first year in college, women are no less likely than men to complete a degree in a STEM field, including physical sciences and engineering. Further, while student-teacher gender matching in middle and high school math and science courses appeared to boost freshman-year STEM course taking for women, the same is not true for completion of a STEM major conditional on first-year coursework. In fact, the correlation between gender matching of students and teachers in middle and high school and STEM major persistence is frequently negative. Further, none of the student-professor matching variables is positively correlated with the likelihood of completing any degree in STEM or completing a particular STEM major.

D. Decomposition of Completing a College Degree in a STEM Major

In order to gauge the relative importance of the factors that contribute to gender differences in eventual college major completion, I conduct a decomposition analysis in the spirit of Arcidiacono and Koedel (2014). The overall gender gap is the difference in the predicted probabilities of completing a given major for men and for women. These unconditional probabilities are the product of each gender's conditional probabilities at each stage: graduating with a STEM major (conditional on first-year course taking, college entry and all pre-college outcomes), first-year coursework (conditional on entering college and pre-college outcomes), college entry (conditional on earning a high school diploma and pre-high-school test scores) and high school graduation (conditional on test scores in 5^{th} and 8^{th} grade).

The portion of overall gap that is attributable to gender differences at each stage can be determined by assigning women the relevant values for men (while keeping the values for women at prior stages constant) and recalculating the differences in predicted STEM major attainment probabilities between men and women. Details are provided in the Appendix.

Results of the decomposition analysis are presented in Table 13. While the absolute differences in the predicted probability of STEM major completion may seem small, it is important to recognize that even for males, the predicted likelihood of graduating from high school, attending college and completing a STEM major is less than five percent.

For both life sciences (biology) and physical sciences (chemistry and physics) a large share of the overall STEM major completion gap is attributable to math achievement differentials in fifth grade. Conditional on fifth-grade scores, eighth-grade scores have a much smaller contribution and in fact tend to reduce the STEM major gap. This is consistent with the improvement in math test scores for girls, relative to boys, between the end of elementary and the end of middle school (Table 6).

The gender differences in high school graduation and college entry have little impact on STEM major completion. Likewise, first-year college coursework decisions have modest impacts, though the effects are more substantial in the physical sciences. For STEM majors as a whole, as well as sub-disciplines in physical and life sciences, there is still a substantial proportion of the gender gap in major completion that is unexplained by student achievement prior to high school, high school completion and college entry or by the choice of courses in the freshman year in college.

V. Summary and Conclusions

Growing concern about the low production of college graduates in STEM fields, particularly among minorities and women, has led to a rapidly growing research literature seeking to understand the causes of these disparities and hence provide guidance as to appropriate policies. The focus of this research has been on the experiences of students once they attend college, including the identity of their instructors and the institutions they enroll in. This college-level focus forecloses the possibility that pre-college experiences, such as the quality and identity of middle and high school instructors and peers can shape future major choices in college.

In this paper, I present new data tracking individual students from elementary school through college. Gender gaps in math achievement are generally modest throughout elementary, middle and high school and women are more likely to successfully complete high school and attend college. Once they get to college, however, they are much less likely than males to obtain a bachelor's degree in a STEM field. Although female college students are more likely to complete a major in biology or other life-science fields, they are much less likely than men to earn a degree in engineering or the physical sciences. Exposure to female math and science teachers in middle and high school is correlated with increases in the number of STEM courses taken by female college freshmen. Likewise, students whose middle and high school math and science teachers held degrees in the relevant field, rather than in education, were more likely to take STEM courses as college freshman. However, the gender matching of students and teachers in secondary school does not increase persistence of women in STEM fields after their freshman year. Similarly, conditional on first-year coursework, exposure to female instructors in STEM courses taken during the freshman year does not increase the likelihood of successfully completing a major in a STEM field.

These findings have several important implications for policy and for future research. First, it is important to realize that underrepresentation of women is not uniform across STEM fields; while women constitute a disproportionally low share of engineering, math and physical science graduates, they have a higher likelihood than men of obtaining a degree in the biological sciences (conditional on attending college). Thus, policies designed to enhance the presence of women should focus on engineering and physical sciences, rather than STEM more broadly. Second, the results suggest that altering the gender composition of college faculty is unlikely to substantially change the relative numbers of women in STEM fields. Rather, interventions at the middle and high school level appear more likely to influence young women's college course selection and ultimately completion of a STEM major. While manipulating the gender mix of students in secondary math and science classrooms does not appear to increase STEM course taking by women when they enter college, the characteristics of their middle and high school math and science teachers do appear to influence course selection in college. In particular, increasing the proportion of female math and science teachers in high school and hiring more high school math and science teachers with degrees in the relevant subject area appear likely to boost the number of STEM courses taken by women during their first year in college. The increased first-year coursework in turn will tend to boost the numbers of women who successfully complete a STEM major.

References

Anelli, Massimo and Giovanni Peri (2017). The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, Doi: 10.1111/ecoj.12556.

Arcidiacono, Peter and Cory Koedel (2014). Race and College Success: Evidence from Missouri. *American Economic Journal: Applied Economics*, 6(3): 20-57.

Bettinger, Eric (2010). "To Be or Not to Be: Major Choices in Budding Scientists." *American Universities in a Global Market*. Ed. Charles Clotfelter. Chicago: University of Chicago Press, pp. 69-98.

Bettinger, Eric and Bridget Terry Long (2005). Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *American Economic Review*, 95(2), 152-157.

Bottia, Martha Cecilia, Elizabeth Sterns, Roslyn Arlin Mickelson, Stephanie Moller and Lauren Valentino (2015). "Growing the Roots of STEM Majors: Female Math and Science High School Faculty and the Participation of Students in STEM," *Economics of Education Review*, 45, 14-27.

Carrell, Scott E., Marianne E. Page and James E. West (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics*, 125(3), 1101-1144.

Card, David and A. Abigail Payne (2017). High School Choices and the Gender Gap in STEM. Cambridge, MA: NBER Working Paper No. 23769.

Dee, Thomas S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, 42(3), 528-554.

Greene, William (2004). The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects. *Econometrics Journal*, 7, 98-119.

Griffith, Amanda L. (2010). Persistence of Women and Minorities in STEM Field Majors: Is it the School that Matters? *Economics of Education Review*, 29(6), 911-922.

Hoffmann, Florian and Philip Oreopoulos (2009). A Professor Like Me: The Influence of Instructor Gender on College Achievement. *Journal of Human Resources*, 44(2), 479-494.

Kokkelenberg, Edward C., and Eshna Sinha (2010). Who Succeeds in STEM Studies? An Analysis of Binghamton University Undergraduate students. *Economics of Education Review*, 29(6), 935-946.

Park, Hyunjoon, Jere R. Behrman, and Jaesung Choi (2018). Do Single-Sex Schools Enhance Students' STEM (Science, Technology, Engineering, and Mathematics) Outcomes? *Economics of Education Review*, 62, 35-47.

Price, Joshua (2010). The Effect of Instructor Race and Gender on Student Persistence in STEM Fields. *Economics of Education Review*, 29(6), 901-910.

Shi, Ying (2018). The Puzzle of Missing Female Engineers: Academic Preparation, Ability Beliefs, and Preferences. *Economics of Education Review*, 64, 129-143.

Table 1 – Descriptive Statistics by 5th-Grade Cohort
(Students with a test score in grade 5 who are enrolled three or more years later in a public school in Grade 9)

	Cohort			
	1997	**1998**	**1999**	**2000**
Female	0.5000	0.4971	0.4954	0.4982
White	0.5292	0.5366	0.5211	0.5119
Black	0.2469	0.2490	0.2483	0.2409
Hispanic	0.1972	0.1847	0.1996	0.2138
Asian	0.0196	0.0215	0.0214	0.0216
Race-other	0.0072	0.0082	0.0097	0.0118
Free Lunch	0.4253	0.4125	0.4146	0.4082
Reduced-Price Lunch	0.0931	0.1036	0.1048	0.1093
5th-grade SSS Normed Math Score	0.0050	0.0075	0.0151	0.0199
5th-grade NRT Normed Math Score			0.0156	0.0221
6th-grade SSS Normed Math Score	-0.5011	-0.5446	0.0661	0.0809
6th-grade NRT Normed Math Score	-0.5981	0.0721	0.0638	0.0779
7th-grade SSS Normed Math Score	-0.5654	0.0850	0.0735	0.0821
7th-grade NRT Normed Math Score	0.0590	0.0736	0.0695	0.0790
8th-grade SSS Normed Math Score	0.0636	0.0801	0.0688	0.0782
8th-grade NRT Normed Math Score	0.0515	0.0703	0.0564	0.0675
9th-grade SSS Normed Math Score	0.1327	0.1269	0.1233	0.1156
9th-grade NRT Normed Math Score	0.1199	0.1081	0.1045	0.0960
10th-grade SSS Normed Math Score	0.0914	0.2122	0.1828	0.2005
10th-grade NRT Normed Math Score	0.0566	0.0637	0.0551	0.0738
Earned Regular HS Diploma within 4 years of Entering Grade 9	0.6299	0.6095	0.6213	0.6429
Earned Regular HS Diploma within 5 years of Entering Grade 9	0.6323	0.6121	0.6247	0.6465

Table 2 - Descriptive Statistics by 5th-Grade Cohort
(Students with a test score in grade 5 who are enrolled exactly four years later in a public school in Grade 9)

	Cohort			
	1997	**1998**	**1999**	**2000**
Female	0.5167	0.5127	0.5111	0.5132
White	0.5408	0.5481	0.5334	0.5249
Black	0.2327	0.2354	0.2344	0.2267
Hispanic	0.1985	0.1856	0.1999	0.2138
Asian	0.0211	0.023	0.0229	0.0231
Race-other	0.0069	0.008	0.0095	0.0114
Free Lunch	0.3985	0.3883	0.3896	0.384
Reduced-Price Lunch	0.0927	0.1035	0.1048	0.1091
5th-grade SSS Normed Math Score	0.0881	0.0797	0.0851	0.0843
5th-grade NRT Normed Math Score			0.0857	0.0925
6th-grade SSS Normed Math Score	-0.2911	-0.5562	0.1345	0.1462
6th-grade NRT Normed Math Score	-0.6128	0.1426	0.133	0.1457
7th-grade SSS Normed Math Score	-0.7875	0.1486	0.1361	0.1449
7th-grade NRT Normed Math Score	0.1309	0.1401	0.1354	0.145
8th-grade SSS Normed Math Score	0.1256	0.1375	0.1296	0.1371
8th-grade NRT Normed Math Score	0.1165	0.1309	0.1185	0.1277
9th-grade SSS Normed Math Score	0.1933	0.1797	0.1768	0.1682
9th-grade NRT Normed Math Score	0.1798	0.1628	0.1581	0.1501
10th-grade SSS Normed Math Score	0.1232	0.2457	0.2143	0.2327
10th-grade NRT Normed Math Score	0.0916	0.0956	0.0884	0.1077
Earned Regular HS Diploma within 4 years of Entering Grade 9	0.6799	0.6537	0.6652	0.6833
Earned Regular HS Diploma within 5 years of Entering Grade 9	0.6823	0.6563	0.6686	0.6867

Table 3 - Normed Test Scores in Grades 5-8 by Race/Ethnicity and Gender (Students with a test score in grade 5 in 2000 who are enrolled four years later in a public school in Grade 9)

Race/Ethicity and Gender		Math Test Scores							
		SSS Grade 5	NRT Grade 5	SSS Grade 6	NRT Grade 6	SSS Grade 7	NRT Grade 7	SSS Grade 8	NRT Grade 8
White	- All	0.3186	0.3308	0.3674	0.4028	0.3633	0.3868	0.3413	0.3731
	- Male	0.3585	0.3479	0.3786	0.4308	0.3848	0.4139	0.3325	0.3829
	- Female	0.2792	0.3139	0.3564	0.3753	0.3423	0.3603	0.3499	0.3636
Black	- All	-0.4070	-0.3815	-0.3276	-0.3609	-0.3344	-0.3611	-0.3532	-0.3879
	- Male	-0.4602	-0.4532	-0.4016	-0.4047	-0.4116	-0.4048	-0.4441	-0.4383
	- Female	-0.3609	-0.3195	-0.2647	-0.3236	-0.2692	-0.3244	-0.2768	-0.3456
Hispanic	- All	-0.0230	-0.0502	0.0400	-0.0186	0.0440	0.0048	0.0762	-0.0041
	- Male	-0.0143	-0.0627	0.0243	-0.0117	0.0339	0.0112	0.0552	-0.0012
	- Female	-0.0311	-0.0385	0.0546	-0.0250	0.0534	-0.0012	0.0957	-0.0069
Asian	- All	0.5680	0.6522	0.7132	0.7832	0.7268	0.8221	0.7659	0.7749
	- Male	0.5889	0.6682	0.7156	0.8094	0.7414	0.8392	0.7747	0.8038
	- Female	0.5481	0.6370	0.7110	0.7585	0.7129	0.8057	0.7575	0.7476
Other	- All	0.0680	0.0895	0.1730	0.1600	0.1524	0.1408	0.1525	0.1307
	- Male	0.1453	0.1560	0.2252	0.2487	0.2170	0.2152	0.1812	0.1769
	- Female	-0.0090	0.0234	0.1219	0.0725	0.0883	0.0662	0.1239	0.0843
All	- Male	0.1060	0.0931	0.1425	0.1641	0.1482	0.1630	0.1183	0.1345
	- Female	0.0639	0.0919	0.1496	0.1284	0.1418	0.1283	0.1546	0.1214

Table 4 – High School Exit by Race/Ethnicity and Gender
(Students with a test score in grade 5 in 1997-2000 who are enrolled four years later in a public school in Grade 9)

Race/Ethnicity and Gender	No Exit (Enrolled in Each of Grades 9-12)	Dropped out	Exit to home school	Exit to private school	Exit – other
White - All	188,875 [69.67]	15,142 [5.59]	4,619 [1.70]	5,429 [2.00]	57,050 [21.04]
- Male	91,445 [67.86]	8,275 [6.14]	1,983 [1.47]	2,689 [2.00]	30,372 [22.54]
- Female	97,430 [71.46]	6,867 [5.04]	2,636 [1.93]	2,740 [2.01]	26,678 [19.57]
Black - All	73,144 [62.31]	12,571 [10.71]	350 [0.30]	2,174 [1.85]	29,153 [24.83]
- Male	31,339 [57.31]	6,532 [11.94]	168 [0.31]	1,165 [2.13]	15,481 [28.31]
- Female	41,805 [66.67]	6,039 [9.63]	182 [0.29]	1,009 [1.61]	13,672 [21.8]
Hispanic - All	65,797 [65.14]	9,874 [9.78]	650 [0.64]	2,649 [2.62]	22,041 [21.82]
- Male	30,466 [62.46]	5,234 [10.73]	261 [0.54]	1,351 [2.77]	11,462 [23.5]
- Female	35,331 [67.64]	4,640 [8.88]	389 [0.74]	1,298 [2.48]	10,579 [20.25]
Asian - All	9,585 [83.76]	312 [2.73]	51 [0.45]	107 [0.93]	1,389 [12.14]
- Male	4,686 [82.36]	171 [3.01]	18 [0.32]	51 [0.9]	764 [13.43]
- Female	4,899 [85.14]	141 [2.45]	33 [0.57]	56 [0.97]	625 [10.86]
Other - All	2,725 [59.05]	388 [8.41]	78 [1.69]	110 [2.38]	1,314 [28.47]
- Male	1,335 [58.58]	190 [8.34]	34 [1.49]	60 [2.63]	660 [28.96]
- Female	1,390 [59.5]	198 [8.48]	44 [1.88]	50 [2.14]	654 [28]
All - Male	162,205 [64.87]	20,590 [8.23]	2,488 [1.00]	5,380 [2.15]	59,381 [23.75]
- Female	184,154 [69.87]	18,038 [6.84]	3,323 [1.26]	5,226 [1.98]	52,832 [20.04]

Note: numbers in brackets are row percentages.

Table 5 – Regular High School Diploma Receipt within 5 Years of Entering Grade 9 by Race/Ethnicity and Gender
(Students with a test score in grade 5 in 1997-2000 who are enrolled four years later in a public school in Grade 9)

Race/Ethnicity and Gender		Did Not Receive Diploma	Received Diploma
White	- All	80,914 [29.84]	190,201 [70.16]
	- Male	43,540 [32.31]	91,224 [67.69]
	- Female	37,374 [27.41]	98,977 [72.59]
Black	- All	49,033 [41.77]	68,359 [58.23]
	- Male	25,638 [46.88]	29,047 [53.12]
	- Female	23,395 [37.31]	39,312 [62.69]
Hispanic	- All	32,926 [32.60]	68,085 [67.40]
	- Male	17,517 [35.91]	31,257 [64.09]
	- Female	15,409 [29.50]	36,828 [70.50]
Asian	- All	1,682 [14.70]	9,762 [85.30]
	- Male	919 [16.15]	4,771 [83.85]
	- Female	763 [13.26]	4,991 [86.74]
Other	- All	1,921 [41.63]	2,694 [58.37]
	- Male	967 [42.43]	1,312 [57.57]
	- Female	954 [40.84]	1,382 [59.16]
All	- Male	89,301 [35.71]	160,743 [64.29]
	- Female	78,625 [29.83]	184,948 [70.17]

Note: numbers in brackets are row percentages.

Table 6 – Normed Test Scores in Grades 5-10 by Race/Ethnicity and Gender
(Students with a test score in grade 5 in 2000 who are enrolled four years later in a public school in Grade 9 and continue to be enrolled in Grades 10 and 11)

Race/Ethnicity and Gender		Math Test Scores											
		SSS Grade 5	NRT Grade 5	SSS Grade 6	NRT Grade 6	SSS Grade 7	NRT Grade 7	SSS Grade 8	NRT Grade 8	SSS Grade 9	NRT Grade 9	SSS Grade 10	NRT Grade 10
White	- All	0.3692	0.3889	0.4241	0.4636	0.4217	0.4515	0.4029	0.4425	0.4325	0.4237	0.3857	0.3194
	- Male	0.4178	0.4159	0.4446	0.5015	0.4516	0.4867	0.4031	0.4616	0.4749	0.4493	0.4144	0.3326
	- Female	0.3221	0.3628	0.4045	0.4271	0.393	0.4177	0.4027	0.4242	0.3917	0.3993	0.3577	0.3067
Black	- All	-0.3452	-0.3251	-0.2630	-0.3057	-0.2694	-0.3037	-0.2820	-0.3295	-0.2407	-0.2781	-0.1249	-0.3406
	- Male	-0.3896	-0.3880	-0.3260	-0.3403	-0.3361	-0.3380	-0.3613	-0.3706	-0.2708	-0.3056	-0.1718	-0.3607
	- Female	-0.3081	-0.2726	-0.2113	-0.2773	-0.2150	-0.2758	-0.2174	-0.2960	-0.2160	-0.2558	-0.0856	-0.3247
Hispanic	- All	0.0454	0.0160	0.1099	0.0475	0.1161	0.0729	0.1524	0.0669	0.1731	0.1085	0.2253	0.0248
	- Male	0.0649	0.0171	0.106	0.0674	0.1192	0.0918	0.1452	0.0859	0.2114	0.1291	0.2547	0.0454
	- Female	0.0277	0.015	0.1135	0.0295	0.1133	0.0557	0.1590	0.0498	0.1385	0.0899	0.1981	0.0063
Asian	- All	0.5890	0.6791	0.7345	0.8120	0.7518	0.8525	0.7911	0.8026	0.8364	0.9525	0.7264	0.8733
	- Male	0.6156	0.7039	0.7419	0.8487	0.7732	0.8798	0.8068	0.8404	0.8866	1.0162	0.7929	0.9602
	- Female	0.5639	0.6556	0.7275	0.7776	0.7316	0.8265	0.7763	0.7671	0.7889	0.8925	0.664	0.7915
Other	- All	0.1201	0.1532	0.2481	0.2329	0.2232	0.2148	0.2336	0.2050	0.2520	0.2421	0.2572	0.1321
	- Male	0.2222	0.2478	0.3267	0.3497	0.3110	0.3194	0.2911	0.2812	0.3287	0.3332	0.3274	0.1968
	- Female	0.0228	0.0630	0.1742	0.1224	0.1400	0.1148	0.1793	0.1325	0.1790	0.1538	0.1857	0.0694
All	- Male	0.1761	0.1673	0.2179	0.2397	0.2249	0.2407	0.1993	0.2167	0.2738	0.2379	0.2856	0.1444
	- Female	0.1138	0.1427	0.2027	0.1811	0.1966	0.1849	0.2126	0.1798	0.2032	0.1891	0.2409	0.1030

Table 7 – College Attendance in Year Immediately Following Receipt of Regular High School Diploma by Race/Ethnicity and Gender
(Students with a test score in grade 5 in 1997 who are enrolled four years later in a public school in Grade 9 and graduate within 5 years with a regular high school diploma)

Race/Ethnicity and Gender	No College	FL Community College	4-year FL Public University	4-year FL Private College/Univ	4-year College Out of State
White - All	11,165 [31.06]	12,686 [35.29]	9,054 [25.19]	1,065 [2.96]	1,976 [5.50]
- Male	6,008 [35.38]	5,631 [33.16]	3,832 [22.57]	502 [2.96]	1,007 [5.93]
- Female	5,157 [27.19]	7,055 [37.20]	5,222 [27.53]	563 [2.97]	969 [5.11]
Black - All	4,828 [37.86]	3,343 [26.22]	2,540 [19.92]	449 [3.52]	1,592 [12.48]
- Male	2,325 [42.9]	1,267 [23.38]	846 [15.61]	172 [3.17]	809 [14.93]
- Female	2,503 [34.13]	2,076 [28.31]	1,694 [23.10]	277 [3.78]	783 [10.68]
Hispanic - All	3,741 [29.77]	2,496 [19.86]	3,147 [25.04]	423 [3.37]	2,759 [21.96]
- Male	1,919 [33.81]	1,048 [18.46]	1,277 [22.5]	199 [3.51]	1,233 [21.72]
- Female	1,822 [26.44]	1,448 [21.02]	1,870 [27.14]	224 [3.25]	1,526 [22.15]
Asian - All	274 [15.77]	441 [25.37]	797 [45.86]	84 [4.83]	142 [8.17]
- Male	148 [17.56]	205 [24.32]	374 [44.37]	43 [5.10]	73 [8.66]
- Female	126 [14.08]	236 [26.37]	423 [47.26]	41 [4.58]	69 [7.71]
Other - All	156 [41.49]	115 [30.59]	67 [17.82]		30 [7.98]
- Male	78 [45.09]	47 [27.17]	29 [16.76]		17 [9.83]
- Female	78 [38.42]	68 [33.5]	38 [18.72]		13 [6.4]
All - Male	10,610 [35.72]	8,431 [28.38]	6,519 [21.95]	934 [3.14]	934 [3.14]
- Female	9,802 [28.02]	11,176 [31.95]	9,402 [26.88]	1,139 [3.26]	3,460 [9.89]

Note: numbers in brackets are row percentages. Data security agreements prohibit reporting results with cell sizes less than 10.

Table 8 – Probit Estimates of the Determinants of Taking One or More Courses in a STEM Field in the First Year in College
(Students with a test score in grade 5 in 1997-2000 who are enrolled four years later in a public school in Grade 9 and graduate within 5 years with a regular high school diploma and attend a Florida public university within one year of receiving their diploma)

Explanatory Variables	Estimated Marginal Effect	
Female	-0.0359** (0.0028)	-0.0325** (0.0030)
Black	0.0100* (0.0038)	0.0227** (0.0045)
Hispanic	0.0530** (0.0034)	0.0588** (0.0037)
Asian	0.0304** (0.0058)	0.0270** (0.0061)
Other Race	0.0142 (0.0183)	0.0089 (0.0199)
Free Lunch		-0.0084 (0.0046)
Reduced-Price Lunch		-0.0103 (0.0061)
SSS Grade 9		0.0065 (0.0044)
NRT Grade 9		0.0019 (0.0032)
SSS Grade 10		0.0015 (0.0056)
NRT Grade 10		0.0029 (0.0029)
NRT Grade 7		0.0015 (0.0032)
SSS Grade 8		0.0150** (0.0051)
NRT Grade 8		-0.0031 (0.0034)
SSS Grade 5		-0.0134** (0.0036)
Observations	74,528	67,297

Note: Excludes remedial math courses. All models include cohort controls. Reported estimates are marginal effects. Standard errors in parentheses. * significant at the 5% level, ** significant at the 1% level in a two-tailed test.

Table 9 – Probit Estimates of the Determinants of Taking One or More Courses in a STEM Field in the First Year in College (Students with a test score in grade 5 in 1997-2000 who are enrolled four years later in a public school in Grade 9 and graduate high school within 5 years and attend a Florida public university within one year of receiving their diploma)

Explanatory Variables	Any STEM Field	Biology	Chemistry	Computer Science	Engineer-ing	Math	Physics	Statistics
Female	-0.0325** (0.0030)	0.0613** (0.0034)	-0.0048 (0.0034)	-0.0238** (0.0011)	-0.0820** (0.0020)	-0.0514** (0.0037)	-0.0313** (0.0014)	0.0512** (0.0027)
Black	0.0227** (0.0045)	0.0211** (0.0055)	0.1197** (0.0061)	-0.0002 (0.0012)	0.0229** (0.0033)	0.0377** (0.0056)	-0.0032 (0.0020)	-0.0345** (0.0041)
Hispanic	0.0588** (0.0037)	0.0303** (0.0049)	0.1170** (0.0052)	0.0068** (0.0012)	0.0264** (0.0026)	0.0687** (0.0048)	0.0055** (0.0017)	-0.0128** (0.0037)
Asian	0.0270** (0.0061)	0.0752** (0.0081)	0.1612** (0.0084)	0.0024 (0.0016)	0.0076* (0.0035)	0.0178* (0.0078)	0.0098** (0.0027)	0.0123* (0.0061)
Other Race	0.0089 (0.0199)	-0.0228 (0.0229)	0.0401 (0.0251)	0.0064 (0.0065)	0.0046 (0.0116)	0.0293 (0.0242)	-0.0024 (0.0074)	0.0032 (0.0189)
Free Lunch	-0.0084 (0.0046)	-0.0111* (0.0050)	0.0072 (0.0050)	0.0022* (0.0012)	0.0028 (0.0024)	-0.0009 (0.0056)	0.0042* (0.0020)	-0.0153** (0.0041)
Reduced-Price Lunch	-0.0103 (0.0061)	-0.0088 (0.0066)	0.0021 (0.0066)	0.0001 (0.0014)	-0.0005 (0.0030)	0.0011 (0.0073)	0.0000 (0.0024)	-0.0141* (0.0054)
SSS Grade 10	0.0015 (0.0056)	-0.0149* (0.0065)	0.0667** (0.0061)	0.0017 (0.0011)	0.0199** (0.0025)	-0.0031 (0.0068)	0.0156** (0.0018)	0.0264** (0.0051)
NRT Grade 10	0.0029 (0.0029)	-0.0043 (0.0033)	0.0233** (0.0031)	0.0030** (0.0006)	0.0083** (0.0013)	-0.0140** (0.0034)	0.0099** (0.0009)	-0.0017 (0.0026)
SSS Grade 9	0.0065 (0.0044)	-0.0171** (0.0051)	0.0277** (0.0047)	0.0006 (0.0009)	0.0107** (0.0019)	0.0021 (0.0053)	0.0055** (0.0013)	0.0085* (0.0039)
NRT Grade 9	0.0019 (0.0032)	0.0030 (0.0036)	0.0067 (0.0035)	0.0007 (0.0007)	0.0020 (0.0015)	-0.0061 (0.0038)	0.0029** (0.0010)	0.0009 (0.0029)
SSS Grade 8	0.0150** (0.0051)	-0.0035 (0.0059)	0.0197** (0.0056)	0.0005 (0.0010)	0.0035 (0.0024)	-0.0023 (0.0061)	0.0046** (0.0016)	0.0205** (0.0046)
NRT Grade 8	-0.0031 (0.0034)	-0.0079* (0.0039)	-0.0106** (0.0037)	0.0012 (0.0007)	0.0012 (0.0016)	0.0041 (0.0040)	0.0004 (0.0011)	-0.0007 (0.0030)
NRT Grade 7	0.0015 (0.0032)	-0.0087* (0.0036)	-0.0016 (0.0035)	-0.0016* (0.0007)	-0.0029 (0.0015)	-0.0013 (0.0038)	0.0007 (0.0011)	0.0060* (0.0029)
SSS Grade 5	-0.0134** (0.0036)	-0.0054 (0.0041)	-0.0281** (0.0040)	0.0007 (0.0008)	0.0004 (0.0017)	-0.0119** (0.0044)	0.0005 (0.0013)	-0.0049 (0.0033)
Observations	67,297	67,297	67,297	67,297	67,297	67,297	67,297	67,297

Note: Excludes remedial math courses. All models include cohort controls. Reported estimates are marginal effects. Standard errors in parentheses. *significant at the 5% level, **significant at the 1% level in a two-tailed test.

Table 10 – Probit Estimates of the Determinants of Taking >=1 Courses in any STEM Field in the First Year in College (Students with a test score in grade 5 in 1997-2000 who are enrolled four years later in a public school in Grade 9 and graduate high school within 5 years and attend a Florida public university within one year of receiving their diploma)

Explanatory Variables	Probit Without HS Fixed Effects	Linear Probability Without HS Fixed Effects	Linear Probability With HS Fixed Effects	Linear Probability with HS Fixed Effects (Schools with 100+ Teachers)	Probit IV
Female	0.0093 (0.0246)	0.0105 (0.0251)	-0.0330 (0.0255)	-0.0799** (0.0300)	-0.6874** (0.1591)
Female x Proportion of Middle and HS Math and Science Courses Taught by a Female Teacher	0.0607** (0.0145)	0.0623** (0.0148)	-0.0005 (0.0151)	-0.0054 (0.0175)	1.4025** (0.2179)
Enrolled in at Least One HS Biology Course Taught by a Teacher with a BA in Biology	0.0142* (0.0060)	0.0147** (0.0062)	-0.0152* (0.0067)	-0.0107 (0.0076)	-0.1478 (0.0793)
Enrolled in at Least One HS Chemistry Course Taught by a Teacher with a BA in Chemistry	0.0203* (0.0075)	0.0206** (0.0078)	0.0142 (0.0086)	0.0223* (0.0093)	0.0214 (0.0739)
Enrolled in at Least One HS Physics Course Taught by a Teacher with a BA in Physics	-0.0131 (0.0124)	-0.0137 (0.0118)	-0.0163 (0.0132)	-0.0168 (0.0133)	0.3983** (0.1343)
Enrolled in at Least One HS Math Course Taught by a Teacher with a BA in Math	0.0132* (0.0051)	0.0134* (0.0053)	0.0069 (0.0058)	0.0016 (0.0065)	0.1538** (0.0526)
Female x Proportion of Female Students in Middle and HS Math and Science Courses	-0.1467** (0.0405)	-0.1502** (0.0416)	-0.0045 (0.0427)	0.0924 (0.0503)	-0.5112** (0.1600)
Observations	49,633	49,633	49,627	33,230	49,570

Note: Excludes remedial math courses. All models include cohort controls, student race/ethnicity indicators, student race/ethnicity interactions with teacher race/ethnicity, student race/ethnicity interactions with peer student race/ethnicity, controls for free and reduced-price lunch in grade 5 and controls for math test scores in grades 5, 7, 8. Reported estimates are marginal effects. Standard errors in parentheses. *significant at the 5% level, ** significant at the 1% level in a two-tailed test.

Table 11 – Probit Estimates of the Determinants of Earning a Bachelor's Degree in a STEM Major Within 9 years of Starting Grade 9 – *With and Without Pre-College Controls*
(Students with a test score in grade 5 in 1997 who are enrolled four years later in a public school in Grade 9 and graduate within 5 years with a regular high school diploma and attend a Florida public university within one year of receiving their diploma)

Explanatory Variables	Earning a Bachelor's Degree in STEM	Earning a Bachelor's Degree in Bio. Sci.	Earning a Bachelor's Degree in Chemistry	Earning a Bachelor's Degree in Engineering	Earning a Bachelor's Degree in Math	Earning a Bachelor's Degree in Physics
No Pre-College Controls						
Female	-0.0475**	0.0016**	-0.0022**	-0.0412**	-0.0022**	-0.0030**
	(0.0019)	(0.0004)	(0.0005)	(0.0013)	(0.0004)	(0.0004)
Observations	75,292	75,292	75,292	75,292	75,292	62,816
With Pre-College Controls for Family Income and Student Achievement						
Female	-0.0254**	0.0020**	-0.0004	-0.0227**	-0.0005*	-0.0011**
	(0.0017)	(0.0005)	(0.0004)	(0.0011)	(0.0002)	(0.0002)
Observations	67,966	67,966	67,966	67,966	67,966	56,776

Note: All models include cohort controls and controls for race/ethnicity. Estimates in the lower panel are from models that include controls for free and reduced-price lunch in grade 5 and controls for math test scores in grades 5, 7, 8, 9 and 10. Reported estimates are marginal effects. Standard errors are in parentheses. *significant at the 5% level, ** significant at the 1% level in a two-tailed test.

Table 12 – Probit Estimates of the Determinants of Earning a Bachelor's Degree in a STEM Major Within 9 years of Starting Grade 9 (Students with a test score in grade 5 in 1997 who are enrolled four years later in a public school in Grade 9 and graduate high school within 5 years and attend a Florida public university within one year of receiving their diploma)

Explanatory Variables	Bachelor's Degree in any STEM	Bachelor's Degree in Bio. Sci.	Bachelor's Degree in Chemistry	Bachelor's Degree in Engineering	Bachelor's Degree in Math	Bachelor's Degree in Physics
Female	0.0183 (0.0132)	0.0020 (0.0030)	0.0018 (0.0026)	0.0189** (0.0071)	-0.0003 (0.0018)	0.0100** (0.0099)
Female x Proportion of Middle and HS Math and Science Courses Taught by a Female Teacher	-0.0186* (0.0082)	0.0004 (0.0019)	-0.0036* (0.0016)	-0.0072 (0.0039)	-0.0001 (0.0010)	-0.0003 (0.0005)
Enrolled in at Least One HS Biology Course Taught by a Teacher with a BA in Biology	-0.0054 (0.0029)	0.0006 (0.0009)	-0.0003 (0.0006)	-0.0022 (0.0010)	-0.0004 (0.0003)	0.0001 (0.0002)
Enrolled in at Least One HS Chemistry Course Taught by a Teacher with a BA in Chemistry	0.0089* (0.0042)	-0.0002 (0.0010)	-0.0009 (0.0005)	-0.0006 (0.0014)	0.0005 (0.0006)	0.0001 (0.0002)
Enrolled in at Least One HS Physics Course Taught by a Teacher with a BA in Physics	0.0152** (0.0064)	-0.0025 (0.0007)	0.0015 (0.0014)	0.0063** (0.0028)	0.0024* (0.0014)	0.0005 (0.0005)
Enrolled in at Least One HS Math Course Taught by a Teacher with a BA in Math	0.0011 (0.0026)	-0.0008 (0.0006)	0.0009 (0.0006)	-0.0004 (0.0009)	0.0000 (0.0003)	0.0000 (0.0001)
Female x Proportion of Female Students in Middle and HS Math and Science Courses	-0.0367 (0.0230)	-0.0004 (0.0052)	-0.0007 (0.0045)	-0.0459** (0.0113)	0.0004 (0.0028)	-0.0045** (0.0016)
Female x Proportion of First-Year College STEM Courses Taught by a Female Instructor	-0.0156** (0.0042)	-0.0029** (0.0010)	-0.0014 (0.0009)	-0.0059** (0.0021)	-0.0009 (0.0006)	0.0001 (0.0002)
No. of Computer Courses in First Year	0.0228** (0.0022)	-0.0024 (0.0023)	-0.0016 (0.0012)	0.0001 (0.0009)	0.0004** (0.0002)	0.0000 (0.0001)
No. of Engineering Courses in First Year	0.0187** (0.0010)	-0.0011 (0.0005)	-0.0011** (0.0004)	0.0076** (0.0005)	-0.0004* (0.0002)	-0.0001* (0.0000)
No. of Math Courses in First Year	0.0052** (0.0010)	-0.0000 (0.0003)	0.0000 (0.0002)	0.0027** (0.0004)	0.0006** (0.0001)	0.0001 (0.0000)
No. of Statistics Courses in First Year	-0.0277** (0.0025)	-0.0012* (0.0006)	-0.0019** (0.0005)	-0.0129** (0.0013)	0.0001 (0.0002)	-0.0005** (0.0002)
No. of Science Courses in First Year	0.0113** (0.0005)	0.0010** (0.0001)	0.0010** (0.0001)	0.0015** (0.0002)	0.0000 (0.0001)	0.0001** (0.0000)

Observations	48,550	48,281	48,550	48,550	48,550	39,867

Note: All models include cohort controls, controls for free and reduced-price lunch in grade 5 and controls for math test scores in grades 5, 7, and 8. Reported estimates are marginal effects. Standard errors are in parentheses. *significant at the 5% level, ** significant at the 1% level in a two-tailed test. Excludes remedial math courses. All models include cohort controls.

Table 13 – Decomposition of Male-Female STEM Degree Completion Gap

Degree	Predicted Degree Completion Probability for Males (Percent)	Predicted Degree Completion Probability for Females (Percent)	Male-Female Gap in Predicted Degree Completion Probability (Percentage Points)	Percentage of Predicted Gap Explained by:					
				Grade 5 Math Test Scores	Grade 8 Math Test Scores	High School Graduation	College Entry	First-Year College Course-work	Other Factors
Any STEM	.04672	.02844	0.01828	73.53	-13.81	1.05	3.75	0.79	34.68
Chemistry	.00076	.00050	0.00026	83.49	-15.68	1.20	4.26	9.15	17.59
Physics	.00005	.00000	0.00004	31.16	-5.85	0.45	1.59	15.21	57.45
Biology	.00052	.00021	0.00031	47.86	-8.99	0.69	2.44	-4.33	62.33

Appendix

Following Arcidiacono and Koedel (AK), the de-composition involves a multi-step process. We begin with an analog of AK's equation (1), which represents the probability of graduating from college with a STEM major:

$$\sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x)\times\Pr(f|c,d,h,x,g)\times\Pr(\,c\,|d,h,x,g)$$

$$\times\Pr(d|h,x,g)\times\Pr(h|x,g)\times\Pr(x|g) \qquad [1]$$

where

Y= graduate from college with a STEM major
f = first-year coursework
c =enroll in college
d = student earns a high school diploma
h = vector of high school characteristics (e.g. peer composition)
x = pre-HS academic background (measured by test scores)
g =gender

The overall difference in graduation rates between males and females (D_g) therefore equals:

$$\begin{aligned} D_g = {} & \textstyle\sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x,male)\times\Pr(f|c,d,h,x,male)\times \\ & \Pr(\,c\,|d,h,x,male)\times\Pr(d|h,x,male)\times\Pr(h|x,male)\times\Pr(x|male)- \\ & \textstyle\sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x,female)\times\Pr(f|c,d,h,x,female)\times \\ & \Pr(\,c\,|d,h,x,female)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female) \end{aligned}$$

The first step in the decomposition is to determine how that difference in predicted probabilities would change if women had the same conditional stem graduation rate as men, but their own true values for all of the other components (f, c, d, h, x):

$$D_f = \sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x,male)\times\Pr(f|c,d,h,x,female)\times \Pr(\,c\,|d,h,x,female)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female) - \sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x,female)\times\Pr(f|c,d,h,x,female)\times \Pr(\,c\,|d,h,x,female)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female)$$

The next step in the decomposition is to determine how that difference in predicted probabilities would change if women also had the same first-year college coursework as men, but their own true values for all of the other components (c, d, h, x):

$$D_f = \sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x)\times\Pr(f|c,d,h,x,male)\times \Pr(\,c\,|d,h,x,female)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female) - \sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x)\times\Pr(f|c,d,h,x,female)\times \Pr(\,c\,|d,h,x,female)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female)$$

The third step is to determine how the gap would change further if women had both the same first-year coursework as men and the same probability of attending college:

$$D_f = \sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x)\times\Pr(f|c,d,h,x,male)\times \Pr(\,c\,|d,h,x,male)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female) - \sum_{x\in X}\sum_{h\in H}\sum_{d\in D}\sum_{c\in C}\sum_{f\in F}\Pr(Y=1|f,c,d,h,x)\times\Pr(f|c,d,h,x,male)\times \Pr(\,c\,|d,h,x,female)\times\Pr(d|h,x,female)\times\Pr(h|x,female)\times\Pr(x|female)$$

The effects of earning a high school diploma, high school characteristics and pre-high-school achievement are determining by continuing in a similar recursive fashion.